Grenze International Journal of Engineering and Technology, June Issue

A Comprehensive Review of Generative Adversarial Network (GAN) based Image Inpainting

Tammineni Shanmukhaprasanthi[1], Swarajya Madhuri Rayavarapu[2], Yenneti Laxmi Lavanya[3] and Gottapu Sasibhushana Rao[4]
[1-4]AUCE, Andhra University, Visakhapatnam, India
Email: prashanthitammineni.rs@andhrauniversity.edu.in, madhurirayavarapu.rs@andhrauniversity.edu.in, lavanya9234@gmail.com, sasigps@gmail.com

***Abstract*—Image inpainting is a method that uses AI algorithms to recreate missing picture parts in order to "fix" or "correct" an image. That's because it "fixes" or "corrects" the original photo. As image inpainting techniques have progressed and gotten more complex, the significance of Image Inpainting algorithms has grown in recent years. It may be used to edit images in several ways, including erasing unwanted elements, reducing background noise, and adding missing details. To do this, it is required to present an overview of picture inpainting techniques that may serve as a thorough introduction to the field. Although there are insufficient existing resources that offer a full review of all the image Inpainting techniques that are now in use, this investigation was conducted. As the area of image inpainting develops, researchers may want documentation detailing the many image inpainting training techniques available for use while developing their models. In order to aid newcomers in learning more about image Inpainting, this study provides a comprehensive overview of the topic. The major goal of this research was to synthesize existing work on image Inpainting by summarizing the various methods used and the domains in which they were used.**

***Index Terms*— Generative Adversarial Networks, Image Inpainting, Generator, Discriminator.**

I. INTRODUCTION

These days, image inpainting is acknowledged as being among the most exciting new areas of study in computer vision. These methods employ what is currently there in the image to infer and align details about missing or damaged components. This helps in making sure the recreated picture adheres as closely as possible to the standards of the human visual perception. In the Medieval Ages, when this technique was first utilized, specialists would use their knowledge and skill to hand repair damaged artwork images. This method made its debut. It has now been adopted for usage in many other kinds of computer vision tasks, such as photo manipulation, movie and TV effects, augmented and virtual reality, and even the protection of digital artifacts.
The Generative Adversarial Network (GAN) is a popular tool for fixing incomplete pictures. Deep Convolutional GANs consist of two parts: a generator and a discriminator. The discriminator's job is to try to tell the difference between the original image and the fixed one created by the generator. The fixed picture must be made by the generator. The generator uses a variety of decoders to up sample the image. Decoders typically consist of a ReLU activation, fractional-strided convolutional layer, a batch normalization layer. In contrast, a transposed convolutional layer plus a Tanh activation are all that make up the final decoder. In contrast, the

Grenze ID: 01.GIJET.9.2.16

discriminator just examines a tiny subset of each set (the original and the generated images. The first down sampler consists of a single convolutional layer and a Leaky ReLU activation, whereas successive down samplers include strided convolutional layers, batch normalization layers, and Leaky ReLU activations. This is due to the fact that additional down samplers are constructed above the original. a)GAN loss function. The goal of GANs is to produce a replica of a probability distribution. As a result, they ought to make use of loss functions that indicate the distance between the distribution of the data generated by the GAN and the distribution of the actual data.

i)Min-max GAN loss

It is also known as standard GAN loss function. The Discriminator tries to maximize this function while the generator tries to minimize it. The objective function of min-max GAN loss function is given in equation 1.

$$F(x)=[log(D(x))]E_x+E_z[\log(1\text{-}D(G(z)))] \quad (1)$$

It can be further categorized into two parts that is discriminator loss and generator loss.

Discriminator loss

$$F(x) = \nabla\theta_d \frac{1}{m}\sum_{i=1}^{m}\left[\log D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^i\right)\right)\right)\right] \quad (2)$$

In the above equation (2), first term represents probability of generator and maximizing the second term gives fake image that comes from the generator.

Generator loss

$$F(x) = \nabla\theta_d \frac{1}{m}\sum_{i=1}^{m}\log\left(1 - D\left(G\left(z^i\right)\right)\right) \quad (3)$$

This function (given in equation 3) is going to be optimized as much as possible by the generator. In other words, it strives to achieve the highest possible output from the discriminator for its false instances.

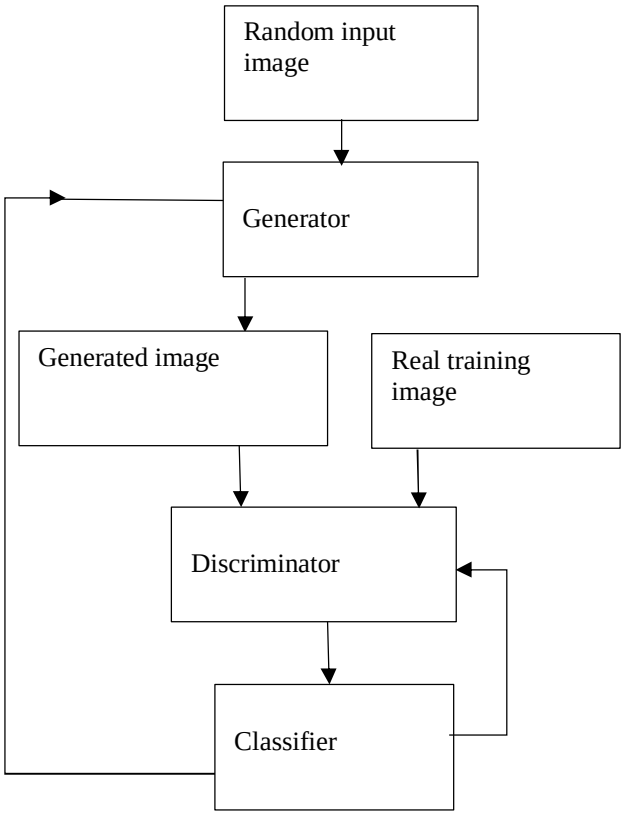

Figure.1 Architecture of Generative Adversarial Network (GAN)

In Section 2, of this study the generative adversarial networks-based image inpainting methods are introduced, including three categories of GAN-based image inpainting methods; and in section 3 challenges and difficulties with GAN is summarized. Section 4 describes the typically used evaluation metrics for image inpainting; and Section 5 concludes with a summary of the image inpainting work and a discussion of future research directions and advances.

II. GAN BASED IMAGE INPAINTING

The process of "deep learning" has only been around for a few years. Utilising the artificial neural networks(ANN's) as the multi-layer neural network (NN) structures and the architecture, it was made to learn different facets of data. It has impressive performance in both visual structure analysis as well as the high-level semantic extraction. Since its introduction in 2014, the Generative Adversarial Networks (GAN) [1] method has become the de facto standard in the field of image restoration. GAN's central idea is to build a set of network structures that, when integrated, can create image textures with the help of the discriminators as well as generators that can tell the difference between the real and fake images.
The discriminator's job is to find better ways to tell the training samples apart, while the generator's job is to narrow the gap between the two types of image distributions. Pathak et al. proposed the use of Neural networks of the Context-Encoders (CE) for the inpainting approach in 2016 [2]. The core idea of the CE is to combine the network architecture of such an encoder-decoder system with that of a structure of a GAN system. With the advent of CE, several refined iterations of CE-based picture inpainting techniques have arisen. These techniques may be broken down into three main groups: enhanced GAN network structure, refined convolutional methods, and refined approaches to integrating high-level and low-level semantics. Since CE's inception, several refined picture inpainting methods that rely on the concept have appeared.

A. An Enhanced GAN Network Architecture for image inpainting.

Pixel continuity, which is correlated to the image's semantic continuity, was previously largely ignored by Image Inpainting methods. This resulted in several color interruptions and line breaks throughout the procedure. Reference [3] is based on the principle of semantic continuity, and it divides the network architecture into two stages: one is fine inpainting and other is rough inpainting. The U-net network is utilized for the first stages of inpainting. This network requires little time to train and has good inpainting outcomes in terms of image composition. The fine inpainting network divides each convolutional layer into two stages and adds a CSA layer i.e., the Coherent Semantic Attention layer to the convolutional layer architecture. The CSA layer is responsible for completing the missing area in the top left corner of the picture. It does this by searching the image for comparable blocks that are complete. After that, the remaining blanks are filled in using the mathematical formula for derivation that was provided, which involves deriving the subsequent pattern from the pattern that came before it. We fix the problem of previous deep learning algorithms' inability to properly reconstruct the original structure in [4]. This is worrisome since it may provide results that are either too smooth or too fuzzy.

B. Technique of Image Inpainting Using an Enhanced Version of the Convolution Method.

For the purpose of ensuring that the produced picture is both worldwide and locally consistent, Reference [5] suggests switching out the discriminator that was included in the first version of the CE model for a pair of different discriminators: global and local. The notion of partial convolution is presented in this study as a means of inpainting irregular areas while taking into consideration only pixels that are helpful. And in the end, the outcome reveals that it is also capable of neatly combining the repaired area with the remainder of the picture even if no further processing was performed afterwards. In this study, we use the GAN-inspired enlarged (CNN) convolutional network structure as a stepping stone. The GAN network structure is also kept as a two-part system, with the generator responsible for creating new pictures and the discriminator checking to see whether they match the originals. In contrast to the Local discriminator, which analyzes the original picture that is positioned on the area which was filled through, the whole image was taken by the global discriminator into account when making its determination of the scene's global consistency On the contrary, The Analysis of the scene's local consistency based on the input picture done by the Global discriminator. The network uses two separate discriminators to guarantee an accurate global observation and improve the quality of the picture fill by adjusting the finer details. Nevertheless, if the mask is large enough to obscure a large number of uniformly spaced objects, the recovery effect will be dampened.
Reference [6] explains how partial convolution may be used to enhance image inpainting. This method has the potential to repair off-center, irregular regions. The analysis of lost photographs with rectangular loss zones that

needed post-processing was the primary focus of early deep learning systems. This study introduces a revolutionary Image Inpainting model, one that can maintain its accuracy even when confronted with holes of varying sizes and shapes. The model infuses inpainting into images by stacking partial convolution processes and updating the mask simultaneously. The network is built similarly to the U-Net design, but all of the convolutional layers which are original have been replaced with convolutional layers which are much partial, and nearest neighbor up-sampling is used in the decoding stage. As the error approaches the image's edge, the partial convolution layer is processed directly rather than convolving using the fill technique, as was mentioned earlier in the article. This is done in place of the convoluted fill method. Incorrect values outside the picture's borders won't be able to affect the repaired border content in any way if this is done.

C. Method for Integrating of the High Level semantics as well as Low Level semantics in Picture Inpainting

Adding to a high-resolution image using a multiscale scale-by-scale recovery approach is described in [7]. This method starts at a less resolution and gradually increases it to a high one. In this article, we distinguish between a texture generation network and content generation network as subsets of the generator network. The missing region is first initialized using the contextual encoder's output, and then high-frequency textures are transmitted in from the boundaries using the contextual encoder's output. We use the content generation network directly to generate pictures that infer the probable contents of the missing region of the image, and we restrict the Image Inpainting process to a predetermined central area. The texture of the content network's output is enhanced by the texture generation network, which is fed the created complementary picture as well as the authentic missing-free image before computing the loss on a given layer of the feature map. To do this, the created complementary picture is combined with the authentic missing-free image and sent into the texture generating network. Knowledge about the patch match is obtained from the similar one that is relatively nearest to the one which is missing, and the calculation of perceptive loss is made more complex. After the matching patches have been identified, the degree of matching is established by altering the amplification value in the network responsible for producing the texture, as opposed to just determining the disparity in pixel values.

As mentioned in Reference [8] by Yu jin Song et al. in ECCV 2018, the inpainting approach may be split into two phases: inference and translation. A network of Image to feature will be taught during the inference stage. This network's tasks include providing initial coarse predictions for the holes and then extracting the holes' characteristics. The projections may be off, but they fill in important details of the underlying structure. A Feature-to-Image network is trained in the translation phase to reconstruct the image from its constituent parts. It makes the missing pieces fit together better and results in an image with detailed, realistic textures. Required to train the Feature-to-Image network may be streamlined with the use of a "patch-swap" layer, which transfers high-frequency texture information from the image's borders to the image's blank spots. This facilitates a simpler training procedure. The closest-looking border image block is used to replace each missing image block in the region. The generated feature map is then used as the input to a Feature-to-Image network. Due to the information included on the boundaries, the feature map provides enough detail to enable high-resolution Picture Inpainting.

Reference [9] makes use of an encoder-decoder structure in a similar fashion. This structure combines lower-order pixel recognition with higher-order semantic recognition in order to guarantee consistency between newly created and previously stored pixels

III. Challenges and Difficulties

Challenges and issues to work through Whilst also the deep learning (DL) has had a lot of success in other areas, its use for Picture Inpainting is only getting started. Many problems, including the following, call for more investigation.

1)The approach for recovering the picture itself [10] relies only on data pertaining to the values of the pixels that need to be restored; this method does not take into account the semantics of the image as a whole.

2)A universal inpainting network is required, and as a result, the universality of inpainting results has to be enhanced [11]. The majority of the findings from the current Image Inpainting research are for images that have very specialized structures.

3) A transition in the rendering process that drastically changes the contour of the object might cause an error in the structure of the picture [12]. All of these problems boil down to one of understanding images, hence finding a solution to understanding the recovered objects' high-level semantic information should be the top priority [13]. This involves identifying the image's context and identifying the item to which the missing part should be added.

There are still a few challenges that regularly come up in real-world implementations of GANs, despite the fact that these networks are quite beneficial in specific application fields. Non-convergence, instability and finally mode collapse are high sensitivity to the hyperparameters are the three main difficulties of GANs. The underlying issue is mode collapse. The collapse of modes is the most basic problem as well as metrics for assessing success. The fact that the generator was operating in collapse mode meant that it could only produce a limited number of typical samples. The fundamental cause of the problems with non-convergence and uncertainty that have been noted is the unbalanced training that GAN has been given.
We offer three methods for reducing the effects of these problems: the design of a new network architecture, the implementation of improved loss functions, and the development of alternative optimization techniques.
The problem of determining which values for the GAN hyperparameters are best may potentially be solved by doing the experiment analysis using weights and biases.

IV. Evaluation Metrics

The performance metrics used are PSNR and SSIM.
PSNR(Peak Signal to Noise Ratio) is the most frequent and widely used picture quality assessment index. It is based on the error that occurs between neighboring pixels, making it an error-sensitive image quality measurement. When the PSNR is higher, the picture may be returned to its original state with more precision. The formula that we use to compute PSNR is as follows: where is the mean square error that occurs between the corrected picture and the ground truth image.

$$PSNR=10\log_{10}\left(\frac{R^2}{MSE}\right) \quad (4)$$

The Structural Similarity Index (SSIM) is a metric that determines how similar two photographs are to one another based on brightness, contrast, and structure. [0,1] is the value range that SSIM covers. When the SSIM value is higher, it indicates that the photos are more similar to one another. Where μ_x and μ_y represent the photographs' respective average gray levels, as seen in the photographs X and Y. The letters σ_x and σ_y stand for the standard deviation (sd) of images X and Y respectively. Covariance between pictures X and Y is denoted by the value σ_{xy}. Both C1 and C2 are infinitesimally small, hence the denominator can never be equal to zero.

$$SSIM(x,y) = \frac{(2\cdot\mu_x\mu_y + c_1).(2.\sigma_{xy} + c_2)}{({c_1}^2+{\mu_x}^2+{\mu_y}^2).({c_2}^2+{\sigma_x}^2+{\sigma_y}^2)} \quad (5)$$

V. Conclusion

The tremendous success of deep learning techniques has facilitated the quick development of Image Inpainting methods. This paper aims to provide academics with a more comprehensive understanding of recent advances in deep learning by providing a summary and review of existing deep learning-based Image Inpainting methods from a variety of perspectives. Much more research has been done on GAN, and the many GAN designs and training approaches that have been developed to combat these challenges are addressed in previous portions of this article. In this article, the development of GAN-based Image Inpainting algorithms and the original architecture of GAN are analyzed and discussed. The purpose of this research is to provide a broad perspective on recent developments and approaches. On the basis of the newly developed taxonomy, an organizational framework for problem-solving is presented. This framework is one that investigators may continue to expand in the future. In the end, the goal of this effort is to disseminate these instructive strategies to new researchers so that they are aware of how to get started with their own research.

References

[1] Goodfellow et al., “Generative adversarial nets,” in Advances in neural information processing systems, 2014.
[2] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, “Context encoders: Feature learning by inpainting,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2536–2544.
[3] H. Liu, B. Jiang, Y. Xiao, and C. Yang, “Coherent semantic attention for image inpainting,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 4170–4179.
[4] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, Edgeconnect: Generative image inpainting with adversarial edge learning. 2019.
[5] S. Iizuka, E. Simo-Serra, and H. Ishikawa, “Globally and locally consistent image completion,” ACM Transactions on Graphics (ToG), vol. 36, no. 4, pp. 1–14, 2017.

[6] F. A. Guilin, K. J. Reda, T.-C. Shih, A. Wang, and B. Tao, “Image inpainting for irregular holes using partial convolutions,” in Proceedings of the European Conference on Computer Vision, 2018, pp. 85–100.
[7] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, “High-resolution image inpainting using multi-scale neural patch synthesis,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 6721–6729.
[8] Y. Song et al., “Contextual-based image inpainting: Infer, match, and translate,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[9] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Generative image inpainting with contextual attention,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5505–5514.
[10] C. Guillemot and O. L. Meur, “Image inpainting: Overview and recent advances,” IEEE signal processing magazine, vol. 31, no. 1, pp. 127–144, 2013.
[11] O. Elharrouss and N. Almaadeed, “Somaya Al-Maadeed, and Younes Akbari,” Neural Processing Letters, vol. 51, no. 2, pp. 2007–2028, 2020.
[12] Z. Xu and J. Sun, “Image inpainting by patch propagation using patch sparsity,” IEEE transactions on image processing, vol. 19, no. 5, pp. 1153–1165, 2010.
[13] Y. Liu and C. Shu, “A comparison of image inpainting techniques,” Sixth International Conference on Graphic and Image Processing, vol. 9443, 2015.